Conference Proceedings
Solar World Congress 2015
Daegu, Korea, 08 – 12 November 2015

STATISTICAL MODELS TO PREDICT SOLAR RADIATION AT HIGH RESOLUTIONS

Philippe Lauret[1], Mathieu David[1], Pierre-Julien Trombe[2] and Faly Ramahatana-Andriamasomanana[1]

[1] PIMENT Laboratory/University of La Reunion, Saint-Denis (France)

[2] Technical University of Denmark, Lyngby (Denmark)

Abstract

In this work, we assess the performance of statistical models for intra-hour solar forecasting. More precisely, a linear recursive model and a nonlinear model are used for generating solar irradiance forecasts at temporal resolutions of a few minutes and over multiple horizons. Our approach is applied to forecasting solar irradiance at single sites using the sole historical ground observations of solar irradiance. The benchmarking of the forecasting methods is made at four sites that exhibit different sky conditions.

Keywords: *intra-hour solar forecasting, statistical models*

1. Introduction

The availability of accurate solar forecasts is of great importance for an efficient integration of large shares of solar energy into the electricity grid (Lorenz and Heinemann, 2012). To ensure reliable grid operation, utilities require accurate forecasts at different granularities and different forecasting time horizons (Kostylev and Pavloski, 2011). Depending on the forecast horizon, different input data and forecasting models are appropriate. Statistical models with on-site measured irradiance are adequate for the very short-term time scale ranging from 5 minutes up to 6 hours (Lorenz and Heinemann, 2012).

Today, solar forecasts are essentially produced on an hourly basis. However, there is a broad consensus among energy experts, electric utilities and regulatory authorities on the future need for intra-hour forecasts in order to accommodate the increasing share of photovoltaic power in power systems.

In this work, the focus is placed on lead times from 10 minutes to 3 hours ahead with a granularity of 10 minutes. Consequently, we assess the performance of two statistical models: a linear recursive model ARMA.rls (Ljung and Söderström, 1983) and a nonlinear model ν-SVM based on support vector machine (Smola and Schölkopf, 2004) for intra-hour solar forecasting. The benchmarking of the two methods is made at four sites that exhibit different sky conditions.

2. Data

Four sites are used to evaluate the performance of the forecasting methods. Three of these sites are island sites: Saint-Pierre and Le Tampon (Reunion island), Oahu-Hawaii (Wilcox and Andreas, 2010). The fourth one is located in Las Vegas (Andreas and Stoffel, 2006). The choice of the aforementioned sites aims at testing the different forecasting techniques for different sky conditions. Table 1 gives the details related to the four locations. The data (two years) used to build the models are Global Horizontal Irradiances (GHI) measured at the four stations. Contrary to the ARMA.rls method, the machine learning technique ν-SVM investigated in this work is a supervised learning method, which consists in learning input-output mappings from empirical data (the training dataset). Consequently, data have been divided into training and test datasets. The test dataset (one year) is used to evaluate the performance of the forecasting techniques. Table 1 gives also the mean GHI of the test year used to compute the relative error metrics (see section 5).

doi:10.18086/swc.2015.07.11 Available at http://proceedings.ises.org

Tab. 1: Locations under study

	Las Vegas (NV)	Oahu (HI)	Saint-Pierre (RUN)	Tampon (RUN)
Provider	NREL	NREL	PIMENT	PIMENT
Position	36.06N 115.08W	21.3N 158.1W	21.3S 55.5E	21.3S 55.5E
Elevation	615m	11m	75m	550m
Climate type	Desertic	Tropical	Tropical	Tropical
Period	2011 (Training) 2012 (Testing)	2010 (Training) 2011 (Testing)	2012 (Training) 2013 (Testing)	2012 (Training) 2013 (Testing)
Mean GHI testing period	544 W.m^{-2}	493 W.m^{-2}	538 W.m^{-2}	457 W.m^{-2}

3. Data processing

The Bird clear sky model (Bird and Huldstrom, 1981) is used to remove the daily and annual seasonalities in the global horizontal solar irradiance (GHI) time series then leading to the definition and use of the clear sky index:

$$k^* = G/G_{clsk} \qquad \text{(eq. 1)}$$

where G is the measured global irradiance and G_{clsk} is the output of the Bird clear sky model. With this methodology, the models designed in this work are dedicated to the stochastic part of the global radiation due to cloud cover, leaving the geometric and the deterministic part to be modeled by the clear sky model.

It is also a common practice to filter out the data in order to remove night hours. This choice is justified because during these periods there is obviously no significant solar radiation to generate electricity (i.e. low potential overnight). We chose to apply a filtering criterion based on the solar zenith angle (θ_Z): solar radiation data for which the solar zenith angle is greater than 80° have been removed. In addition, this filtering process allows to discard data with less precision as measurement uncertainties associated with pyranometers are typically much higher than ± 3.0% for SZA > 80°. Notice also that for the sunrise and sunset, the prediction is also very difficult (mainly in mountainous areas) owing to the geographical shield.

4. Forecasting methods

The performance of the linear ARMA.rls and nonlinear ν-SVM models are appraised against two reference models namely persistence and climatologic mean models.

4.1 Reference models

We propose to test our forecasting methods against reference models like persistence and climatology.

The persistence model is expressed as follows:

$$\widehat{k^*}(t+h) = k^*(t) \qquad \text{(eq. 2)}$$

This model assumes that the clear sky index for each forecasting time horizon h only depends on the previous value, which means that the sky conditions remain invariant between time t and time $t+h$.

We also propose the climatological mean model, which is independent of the forecast time horizon (Lorenz

and Heinemann, 2012). More precisely, this model performs a constant forecast of the clear sky index that corresponds to its mean historical value:

$$\widehat{k^*}(t+h) = mean\,(k^*(t))_{training\ period} \qquad \text{(eq. 3)}$$

4.2 ARMA.rls model

The Auto Regressive Moving Average (ARMA) model is a popular technique in the realm of solar forecasting. In particular, it has been extensively studied in renewable energy forecasting and, owing to its parsimony, it has turned out to be a very tough competitor to beat. Applications include, among others forecasting of wind power generation (Pinson, P., 2012), online power forecasting (Bacher et al., 2009) and wave energy flux (Pinson et al., 2012).

An ARMA(*p,q*) model with p AR terms and q MA terms is defined as follows:

$$\widehat{k^*}(t+h) = \theta_0 + \sum_{i=1}^{p} \theta_i k^*(t-i+1) + \varepsilon(t) + \sum_{j=1}^{q} \phi_j \epsilon(t—j) \qquad \text{(eq. 4)}$$

$\epsilon(t)$ is an independent and identically distributed random variable with a zero mean. The vector $\Theta = \left(\theta_0, \theta_1, \cdots, \theta_p, \Phi_{1,}, \cdots, \Phi_q\right)$ contains the set of parameters to be estimated. A classic setting based on a least-squares method and training data can be used to estimate the set of parameters (Chatfield, 2004).

However, in this work, to estimate the model's parameters, we chose a variation of the least squares method, namely the Recursive Least Square (RLS) method (see Ljung and Söderström, 1983 for details of implementation). This method offers the advantage of reducing the computational cost for estimating the model's parameters. In addition, the parameters are updated in real-time as new data become available (no training set is necessary here). This contrasts with more intensive estimation methods operating on a sliding window where the estimation process is being carried out at each time step. Hence, the RLS method is particularly well suited in an operational context where forecasts have to be timely delivered.

Regarding the structure of the ARMA model, the use of the classical ACF (Auto Correlation Function) and PACF (Partial Auto Correlation Function) techniques (Chatfield, 2004) led to the selection of the following orders $p = 6$ and $q = 2$.

4.3 ν-SVM model

The support vector machine (SVM) is part of the kernel based machine learning techniques used in classification tasks and regression problems (Smola and Schölkopf, 2004). The forecasted clear sky index for time horizon *h* is given by Eq. (5):

$$\hat{k}^*(t+h) = \sum_{i=1}^{n} \alpha_i\, k_{rbf}(\mathbf{x}_i, \mathbf{x}_*) + b \qquad \text{(eq. 5)}$$

k_{rbf} denotes the kernel radial basis function $k_{rbf}(x_p, x_q) = \exp\left[-\gamma\left|(x_p - x_q)\right|\right]$ with hyperparameter γ and b a bias parameter. SVM models are generally stated as a kernel-based method. Indeed, it can be shown that, given *n* training samples, the prediction for an input test vector $\mathbf{x}_*$ can be seen in terms of a linear combination of *n* kernel functions; each one centered on a training point $\mathbf{x}_i$. Notice that $\mathbf{x}_i$ and $\mathbf{x}_*$ are vectors that contain the *p* past values of the clear sky index. A variant of the SVM algorithm called ν-SVM allows controlling the amount of kernel functions used in the regression.

The parameter b (or bias parameter) is derived from the preceding equation and some specific conditions (see Smola and Schölkopf, 2004 for details). The coefficients α_i are related to the difference of two Lagrange multipliers, which are the solutions of a quadratic programming (QP) problem (Smola and Schölkopf, 2004).

Unlike artificial neural networks, which are confronted with the problem of local minima, here the problem is strictly convex and the QP problem has a unique solution. In addition, it must be stressed that not all the training patterns participate in the preceding relationship. Indeed, a convenient choice of a cost function i.e. Vapnik's ε-insensitive function (Smola and Schölkopf, 2004) in the QP problem enables to obtain a sparse solution. The latter means that only some of the coefficients α_i will be nonzero. The examples that come with non-vanishing coefficients are called Support Vectors.

The parameters related to the SVM optimization process are a parameter C that controls the trade-off between overfitting and generalization capability of the algorithm, a parameter ν that controls the amount of support vectors used in the regression and the parameter γ of the covariance function that controls the smoothness of the covariance function (Fonseca Junior et al., 2013). These parameters have been optimized through the use of a k-fold cross-validation procedure (Fonseca Junior et al., 2013). In the present study, regarding the implementation of the support vector regression, we used the LibSVM library (Chang et al., 2011).

5. Results

In the realm of the solar forecasting community, the commonly used error metrics are: the Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Bias Error (MBE). Relative counterparts (rRMSE, rMBE and rMAE) are obtained by normalization with the mean GHI of the test period (see Table 1). The interested reader is referred to Lorenz and Heinemann (2012) for the definition of these error metrics.

In this work, a special focus is placed on the rRMSE. This error metric tends (unlike the rMAE) to be influenced by some extreme events or outliers. Nonetheless, most utility users find this metric suitable as large forecast error results in high financial losses (Lorenz and Heinemann, 2012).

Tables 2, 3 and 4 list respectively the rRMSE, rMBE and rMAE for each forecasting time horizon and for each location. As shown by these tables, one can a priori state that the performances of the different models depend heavily on the sky conditions experienced by each site. The persistence model is a good indicator of the type of climate experienced by a particular site. As seen, it seems that the Las Vegas location experiences high occurrences of clear sky situations while the insular sites (particularly Le Tampon and Oahu) exhibit more variable sky conditions.

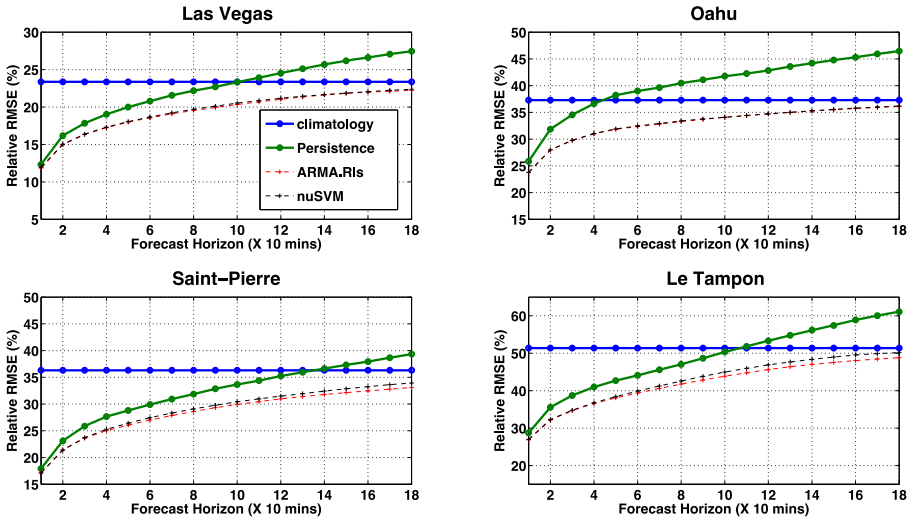

Fig. 1: Accuracy of intra-hour solar forecasting

Fig. 1 plots the rRMSE of the different methods for forecast horizon ranging from 10 minutes to 180 minutes. Fig.1 demonstrates the better performance of the SVM method (ν-SVM) and ARMA.rls model over the persistence model when the forecast horizon increases. As shown by Fig.1, the linear recursive model ARMA.rls performs equally well than the nonlinear SVM model. It even produces slightly better forecasts in the case of Saint-Pierre and Le Tampon. Further, the improvement is more pronounced as the forecast horizon increases. One may notice also that the performances of the linear and nonlinear models tend

towards that of the climatological mean. This behavior is consistent, as these methods tend to asymptotically model the mean of the data.

Overall, contrary to the ν-SVM model, it appears that the ARMA.rls method produces less biased forecasts (see Table 3). Regarding the rMAE metric, except for Las Vegas, the ARMA technique performs better than the SVM method.

Tab. 2: Relative RMSE in % for each location and for each forecast horizon. Pers, RLS and SVM correspond respectively for Persistence, ARMA.rls and ν-SVM models.

h+xx mins	Las Vegas (NV)			Oahu (HI)			Saint-Pierre (RUN)			Tampon (RUN)		
	Climatology		23.4	Climatology		37.3	Climatology		36.3	Climatology		51.3
	Pers	RLS	SVM	Pers	RLS	SVM	Pers	RLS	SVM	Pers	RLS	SVM
h+10	12.3	11.8	11.9	25.8	23.8	23.8	17.9	17.1	17.1	28.7	26.9	27.0
h+20	16.2	15.0	15.0	31.9	28.0	28.0	23.1	21.4	21.4	35.6	32.1	32.3
h+30	17.8	16.3	16.4	34.5	29.7	29.8	25.9	23.6	23.7	38.7	34.7	34.8
h+40	19.0	17.2	17.3	36.6	31.0	31.0	27.6	25.0	25.3	41.0	36.6	36.7
h+50	20.0	18.0	18.0	38.2	31.8	31.9	28.8	26.1	26.4	42.7	38.1	38.4
h+60	20.8	18.6	18.7	39.0	32.3	32.4	29.9	27.0	27.4	44.1	39.4	39.9
h+70	21.6	19.1	19.2	39.6	32.8	32.9	30.9	27.9	28.3	45.6	40.6	41.3
h+80	22.2	19.5	19.7	40.5	33.3	33.4	31.9	28.6	29.0	47.0	41.7	42.6
h+90	22.7	19.9	20.1	41.1	33.7	33.8	32.9	29.3	29.7	48.7	42.8	43.8
h+100	23.3	20.3	20.5	41.7	34.1	34.1	33.6	29.9	30.4	50.3	43.9	45.0
h+110	23.9	20.7	20.8	42.3	34.4	34.4	34.4	30.4	31.0	51.8	44.8	45.9
h+120	24.5	21.0	21.1	42.8	34.7	34.7	35.2	30.1	31.5	53.3	45.6	46.9
h+130	25.1	21.3	21.4	43.6	35.0	35.0	35.9	31.4	31.9	54.8	46.4	47.7
h+140	25.7	21.6	21.7	44.2	35.2	35.3	36.6	31.8	32.4	56.2	47.0	48.4
h+150	26.2	21.8	21.9	44.8	35.5	35.6	37.3	32.1	32.9	57.4	47.5	49.0
h+160	26.6	22.0	22.0	45.3	35.7	35.8	37.9	32.4	33.2	58.9	48.0	49.5
h+170	27.1	22.1	22.2	46.0	35.9	36.0	38.7	32.8	33.6	60.0	48.5	49.9
h+180	27.5	22.2	22.3	46.5	36.1	36.2	39.3	33.1	33.9	61.1	48.8	50.1

Tab. 3: Relative MBE in % for each location and for each forecast horizon

h+xx mins	Las Vegas (NV)			Oahu (HI)			Saint-Pierre (RUN)			Tampon (RUN)		
	Climatology		8.30	Climatology		3.95	Climatology		-4.25	Climatology		1.89
	Pers	RLS	SVM	Pers	RLS	SVM	Pers	RLS	SVM	Pers	RLS	SVM
h+10	0.15	0.15	0.45	0.15	0.47	0.59	0.08	-0.18	-0.65	0.42	0.45	-0.30
h+20	0.27	0.25	1.31	0.30	0.81	1.01	0.15	-0.30	-1.35	0.83	0.71	-0.74
h+30	0.37	0.34	1.79	0.47	1.03	1.23	0.18	-0.37	-2.25	1.24	0.94	-0.80
h+40	0.46	0.37	2.20	0.65	1.21	1.57	0.20	-0.46	-2.52	1.61	1.13	-0.80
h+50	0.56	0.40	2.70	0.82	1.35	1.75	0.20	-0.54	-2.72	1.97	1.29	-0.68
h+60	0.65	0.37	3.06	1.00	1.45	2.20	0.18	-0.64	-2.99	2.30	1.28	-0.37
h+70	0.73	0.36	3.39	1.19	1.52	2.34	0.14	-0.74	-3.26	2.62	1.32	-0.28
h+80	0.80	0.36	3.69	1.37	1.60	2.59	0.10	-0.84	-3.42	2.91	1.33	-0.20

h+90	0.88	0.36	3.93	1.54	1.66	2.56	0.03	-0.92	-3.68	3.18	1.34	0.01
h+100	0.95	0.36	4.11	1.70	1.71	2.55	-0.04	-0.99	-4.07	3.42	1.28	0.19
h+110	0.99	0.35	4.32	1.86	1.72	2.76	-0.14	-1.06	-4.26	3.62	1.21	0.42
h+120	1.04	0.33	4.55	2.00	1.72	2.96	-0.25	-1.15	-4.37	3.79	1.14	0.71
h+130	1.07	0.31	4.78	2.12	1.74	2.87	-0.37	-1.22	-4.50	3.93	1.05	0.95
h+140	1.10	0.29	5.01	2.23	1.76	3.00	-0.50	-1.30	-4.77	4.03	0.89	1.06
h+150	1.12	0.25	5.28	2.33	1.73	3.24	-0.65	-1.39	-4.95	4.10	0.66	1.23
h+160	1.14	0.22	5.48	2.42	1.71	3.39	-0.80	-1.50	-4.96	4.12	0.38	1.34
h+170	1.14	0.20	5.68	2.50	1.68	3.39	-0.96	-1.57	-5.00	4.11	0.17	1.40
h+180	1.13	0.22	5.79	2.56	1.70	3.54	-1.13	-1.71	-5.11	4.07	0.01	1.39

Tab. 4: Relative MAE in % for each location and for each forecast horizon

h+xx mins	Las Vegas (NV)			Oahu (HI)			Saint-Pierre (RUN)			Tampon (RUN)		
	Climatology		14.6	Climatology		29.2	Climatology		29.2	Climatology		42.2
	Pers	RLS	SVM	Pers	RLS	SVM	Pers	RLS	SVM	Pers	RLS	SVM
h+10	5.1	5.5	5.0	16.1	16.1	15.9	9.2	10.1	9.9	16.4	17.6	17.1
h+20	7.1	7.6	6.8	20.5	19.8	19.7	12.4	13.6	14.0	21.5	22.8	22.9
h+30	8.2	8.6	7.7	22.6	21.5	21.5	14.4	15.6	16.7	24.2	25.7	26.1
h+40	9.0	9.3	8.3	24.3	22.7	22.8	15.8	17.1	18.3	26.3	27.7	28.5
h+50	9.7	9.8	8.9	25.6	23.6	23.7	16.9	18.2	19.5	28.1	29.4	30.3
h+60	10.3	10.4	9.4	26.4	24.2	24.4	17.9	19.2	20.6	29.4	30.7	31.9
h+70	10.9	10.9	9.8	27.1	24.7	24.9	18.8	20.5	21.5	30.7	31.9	33.2
h+80	11.5	11.2	10.2	27.9	25.2	25.5	19.7	20.8	22.3	32.1	33.0	34.5
h+90	12.0	11.6	10.5	28.5	25.6	25.9	20.6	21.5	23.0	33.4	34.1	35.7
h+100	12.5	11.9	10.8	29.1	26.1	26.3	21.3	22.1	23.7	34.8	35.0	36.7
h+110	12.9	12.2	11.0	29.7	26.4	26.7	22.1	22.6	24.3	36.2	35.8	37.6
h+120	13.3	12.5	11.2	30.2	26.7	26.9	28.8	23.1	24.8	37.5	36.6	38.4
h+130	13.8	12.7	11.4	30.8	27.0	27.2	23.4	23.5	25.3	38.6	37.4	39.0
h+140	14.2	12.9	11.6	31.3	27.2	27.5	24.1	23.9	25.8	39.9	37.8	39.7
h+150	14.6	13.1	11.7	31.7	27.5	27.7	24.7	24.2	26.3	40.8	38.2	40.2
h+160	14.9	13.3	11.9	32.1	27.7	28.0	25.3	24.5	26.7	42.1	38.8	40.7
h+170	15.3	13.4	12.0	32.7	27.9	28.1	26.0	24.7	27.1	43.1	39.2	41.0
h+180	15.6	13.5	12.1	33.1	28.1	28.3	26.6	25.0	27.4	44.0	39.5	41.2

6. Conclusion

This work proposes a benchmarking of techniques for intra-hour solar forecasting. A machine learning technique such as support vector machine was evaluated against a recursive linear model and reference models like persistence and climatological mean. It was shown that a linear recursive technique like ARMA.rls performs equally well or slightly better than a non-linear method such as support vector machine.

As a conclusion, it appears that a simple technique like ARMA.rls could be a viable solution to predict solar radiation at high resolutions. In addition, the recursive estimation of the model's parameters makes the method very well suited to online forecasting.

7. References

Andreas, A., Stoffel, T., 2006. NREL Report No. DA-5500-56509. University of Nevada (UNLV): Las Vegas, Nevada (Data).

Bacher, P., Madsen, H., Nielsen, H.A, 2009. Online short-term solar power forecasting. Solar Energy. 83, 1772-1783.

Bird, R.E. , Hulstrom, R.L., 1981. Simplified the Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces, Technical Report No. SERI/TR-642-761, Golden, CO: Solar Energy Research Institute.

Chatfield, C., 2004. Time series analysis, an introduction, Chapman & Hall/CRC.

Chang, C.C, Lin, C.J, 2011. LIBSVM: A library for support vector machines, <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

Fonseca Junior, J. G. S., Ozeki, T., Ohtake, H. , Shimose, K., Takashima, T., Ogimoto, K., 2013. Analysis of Different Techniques to Set Support Vector Regression to Forecast Insolation in Tsukuba, Japan. Journal of International Council on Electrical Engineering, 3(2), 121-128.

Kostylev, V., Pavlovski, A., 2011. Solar power forecasting performance towards industry standards. In: Proceedings of the 1st International Workshop on the Integration of Solar Power into Power Systems, October 24, Aarhus, Denmark.

Ljung, L., Söderström, T., 1983. Theory and Practice of Recursive System Identification. Prentice-Hall International

Lorenz, E., Heinemann, D., 2012. Prediction of solar irradiance and photovoltaic power, in: Sayigh, A., (Ed.), Comprehensive Renewable Energy. Elsevier, Oxford, pp. 239-292.

Pinson, P., 2012. Very short term probabilistic forecasting of wind power with generalized logit–normal distributions, Journal of the Royal Statistical Society: Series C (Applied Statistics). 61, 555-576.

Pinson, P., Reikard, G., Bidlot, J.R., 2012. Probabilistic forecasting of the wave energy flux. Applied Energy, 93, 364-370.

Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Statistics and Computing. 14, 199–222

Wilcox, S., Andreas, A., 2010. Solar Resource & Meteorological Assessment Project (SOLRMAP): Rotating Shadowband Radiometer (RSR), Kalaeloa Oahu, Hawaii (Data), NREL Report No. DA-5500-56497.